Journal of Theoretical and Applied Information Technology
15th April 2020. Vol.98. No 07

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# AN INVESTIGATION OF DIGITAL FORENSICS FOR SHAMOON ATTACK BEHAVIOUR IN FOG COMPUTING AND THREAT INTELLIGENCE FOR INCIDENT RESPONSE

**AHMAD K. Al HWAITAT[1], SAHER MANASEER[2], RIZIK M. H. Al-SAYYED[3] , MOHAMMED AMIN ALMAIAH[4], OMAR ALMOMANI[5]**

[12]King Abdullah the II IT School, Department of Computer Science, The University of Jordan, Jordan.

[3]King Abdullah the II IT School, The University of Jordan, Department of Information Technology, Jordan.

[4] King Faisal University, Department computer science, Saudi Arabia.

[5]The World Islamic Sciences and Education University, Department Computer Networks, Jordan

E-mail: [1]Ahmad.Hwaitat1@gmail.com,[2]Sahr@ju.edu.jo , [3]r.alsayyed@ju.edu.jo, malmaiah@kfu.edu.sa[4] , Omar.almomani@wise.edu.jo[5].

**ABSTRACT**

Cyber related crimes are increasing nowadays. Thus digital forensics has been employed in solving cybercrimes. Several researches have been done where they have analysed cyber related attacks, malware types, etc. Researches based on studying and analysing Advanced Persistent Threats (APTs), especially Shamoon attack. This research has been done in order to study and analyse the attacking behaviour of Shamoon malware in fog computing using FPSO (Frequency Particles Swarm Optimization) based on Travelling Salesman approach (TSP). In this proposed system, fog nodes are initiated where the nodes delivers three types of data namely industrial, medical and educational data. Secondly Shamoon attack is created followed by distance matrix evaluation. As the Shamoon attack focuses on attacking industrial data, the attack distribution movement focuses mainly on industrial data. After the evaluation, priorities of the particles should be assigned randomly. Once FPSO parameters are initialized, objective function of every particle is evaluated. The FPSO mechanism implements the working procedure of TSP. Under the FPSO mechanism, swap and insertion operations are performed. In order to find the best shortest path, nearest neighbouring algorithm is used, which follows evaluation of fitness function. After evaluation, local best lbest and global best gbest solutions are obtained. Finally, appropriate positions and velocities are updated. From the resultant optimum path, the distribution of Shamoon attack movement can be analysed. The performance of this proposed system has been evaluated by estimating the fitness value, best cost. The attack distribution of Shamoon data has been observed. Then finally a threat intelligence scheme is proposed for the investigating and analysis behaviour and spread of Shamoon attacks in edges of Fog computing.

***Keywords***: *Fog Computing, Shamoon Attack, FPSO,TSP, Fitness Estimation ,Digital Forensic, Investigation of Cyber Crimes. Cyber Security.*

## 1. INTRODUCTION

Fog Computing [1] [29] is a new archetype, which got extended from Cloud environment by delivering computational resources on the edge of the provided networks. It could be described as same as cloud like environment which has similar computational storage, and application services. However, this fog is decentralized. Additionally, the fog systems has the capability in processing huge amount of local data, quicker operation and are completely portable. It can also be installed on hardware that are heterogeneous in nature. The features of the fog platform are highly suited for location and time sensitive applications [34].

IoT – Internet of Things devices are usually needed for quicker processing of huge amount of data. This massive functionality drives the applications with many security related issues in data transmission and virtualization applications. Malwares or malicious software are becoming sophisticated and comes up with advanced features [2] [35]. They played a part in many of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the computer intrusion incidents and security related incidents. A malware can be considered as a software that causes damage or loss of data to a user, or network that includes Trojan, viruses, spyware, rootkits, etc. Thus malwares are considered as a threat to the future. They are getting worse day by day. Recently several campaign related to malware attacks are targeting middle east countries.

Most of the organizations in Middle East especially, energy, fuel consumption industries are reported continuously that they have been attacked by malwares. Thus it is necessary to mitigate those malwares and intruders. Currently, Shamoon attack [29] [3], which is a famous industrial espionage, played a massive role in destroying Saudi Aramco industrial data. This attack is also referred as W32.DisTrack. It is discovered in the year 2012, where 32 – bit NT kernel types of MS Windows had been targeted. Due to its destructive behaviour and the massive recovery cost, this malware has been remarked to have different behaviour when compared to other malwares. It has three main components namely dropper, wiper and reporter. The Shamoon's components are represented in figure 1.

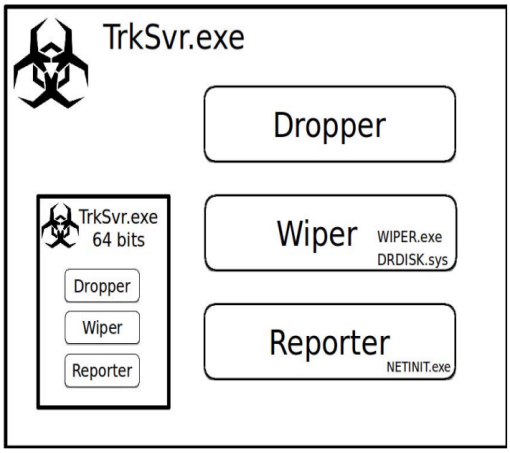


*Figure 1: Shamoon Virus Components*

Thus, it is mandatory to develop an IDS and secured system. In previous phase[30] [31], classification and detection of Shamoon attack has been discussed. In this phase, the analysis of Shamoon attack movement has been discussed using FPSO based on TSP based approach. As TSP [4] is one of the traditional NP-hard problems that has been implemented in various meta heuristic algorithms. For updating best positions and velocities of particles in FPSO, typical TSP steps has been included. The primary objectives of the proposed system are,

- To effectively analyze and study the Shamoon attack using FPSO algorithm based on TSP approach.
- To obtain best predicted features by analyzing the weights (>0.6) using FPSO optimizer.

Generally, the Shamoon attack in Fog computing targets the industrial data. The weaker nodes are spread across the server by tracing the very shortest routeavailable using (TSP).Some of the nodes consists of a loophole which is then exploited by the external attacker to attack the industrial data. So the source of the attack is determined by locating the first node attacked by malware.On finding the first node to be attacked, the source of the attacker can be tracked. The industrial data lost can be recovered by following several techniques. Once the data is recovered, possible evidences present within the data are collected for locating the attacker. With the help of the collected evidences, future attacks can also be prevented. This is the primary aim of this research.

The organization of the remaining paper is as follows: Section 2 explains the related work and literature survey. Section 3 describes the proposed system design along with the details. Section 4 discusses the performance and experimental results. Finally section 5 presents the conclusion of the study.

## 2. THE MATERIAL AND METHOD

Massive extent of problems that can be seen in agriculture, medical, industry, defence, education and other areas like information, can be termed as combinatorial optimization problems [5]. Certain optimization problems come under NP-Hard problems. As these NP-Hard problems were not effectively resolved using conventional methods, several intelligent methodologies had been proposed in the past decades for providing solution namely evolution algorithms like genetic algorithm [6], etc. swarm based intelligence algorithms like PSO[7], BCO[8], etc. This section provides various works related in solving optimization problems, fog computing, digital forensics, cyber-attacks, etc.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### 2.1 Meta-Heuristic algorithms on NP-Hard Problems

A research had been taken place in the field of educational data mining [9], where it explored the efficiency of PSO for a classification purpose i.e. it designed a classification framework that classify the set of questions into 6 levels of Bloom's taxonomy [36]. Thus, it came up with a novel mechanism which is based on an algorithm called RA- Rocchio algorithm in order to eliminate the serious impacts caused by the dimensionality problem in PSO performance. Whenever dealing with multiple (say 1000) features in a dataset, then it is very difficult to attain efficient feature selection. Thus a research [10] had been proposed which utilize binary PSO along with C4.5 as a classification tool. This tool utilized the fitness function for selecting legitimate attributes. Experimentation had been done using 11 datasets and the results were analysed statistically. The paper also concluded that, the BPSO+C4.5 outperformed other classifiers such as Naïve Bayes, SVM, and C4.5.

A learning based technique [11] is used for resolving certain NP-Hard problems. This methodology combined certain deep learning techniques along with useful heuristic elements. A graph based convolutional network is considered as a major component which is utilized for training on how to evaluate the likelihood for each and every vertex in a graph. Here, the network is designed and trained in such a way that it produced a diverse collection of solutions which in turn enables fast exploration through tree based searching. This technique had been tested on four NP-Hard problems and 5 datasets along with benchmark problems.

TSP is considered as one of the NP-Hard problems with greater number of possible solutions. Its complexity would be getting higher along with n factorial for each problem. Thus several heuristic algorithms had been proposed in order to solve TSP problems. [12] focused on six different heuristic algorithms namely Genetic Algorithm, Nearest Neighbour, Tabu Search, ACO, Simulated Annealing, and Tree Physiology based optimization. The best approximation ratio established so far is 3/2 by an algorithm named as Christofides' algorithm which is more than 30 years. The probability of solving TSP is completely based upon algorithms like[36] GA, PSO, ACO, NN, etc. For graphic TSP, this range is further reduced to 13/9.

[13] constructed a novel heuristic methodology known as k-RNN. From the experimentation results, the approximated ratio is 5/4. A paper [14] surveyed and reviewed the heuristics algorithm. These algorithms resolved several problems by providing optimal solutions for smaller size problem and sub-optimal solutions for big problems. This paper also analysed about every strategy used for resolving TSP i.e. time based modification, and its corresponding results. ACO optimization methodology is based on stigmergy nature of insects. ACO provided with promising solutions in dealing TSP related problems. [15] discussed about the alternatives of ACO algorithms and its performance by using a simulation tool termed as TspAntSim. As TSP is one of the NP-Hard problems, it remarkable shown fluctuating behaviour upon the instances from online library called TSPLIB. This paper mainly focused on analysing the behavioural nature of ACO algorithm.

The FPSO optimizer [16] mimics the wave nature of waves by using the features of three parameters namely frequency, wavelength, and amplitude. Here the movement of the particle has been made similar to wave movement. Here, the movement of each and every particle completely depends upon the given frequency. It has been observed that FPSO optimizer improves the primary solution by converging towards the best search space point. Furthermore, this section provides related works about Fog computing, Digital forensics, cyber-attacks.

### 2.2 Fog Computing

Fog computing [17] is commonly known as an extension technology in which it extends cloud technology and the services to the edge of the network. Though, the cloud able to maintain the data processing and traditional demand, it is unsure whether it maintains the demand by the IoT. Thus, fog computing has been providing the best solution in resolving this problem. It also provides services like data, storage, computation and application type services to the users. [18] proposed a fog structured model which is composed of various mobile sinks. Here, the mobile sinks acted as a fog nodes and connected the gap between cloud and WSNs. This paper also designed a parallel algorithm named as DCF which is utilized for attaining low cost schedule, such that to minimize the latency and maximize the throughput. Additionally, it also designed a routing algorithm to maintain energy efficiency, which solves the NP-Hard problem like energy consumption. From experimentation, it had been observed that energy consumption and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

transmission delay were greatly saved and reduced.

### 2.3 Digital Forensics

The progressing number of IoT devices are in need of setting up digital forensic methodologies in order to solve cyber-crimes [19] efficiently. To gain forensic data, it is mandatory for forensic investigators to consider the OS and computing hardware. However, this technique might not be applicable to certain IoT devices. It is very difficult to choose which type of data has to be collected and how the traces can be gained by such investigators. [19] introduced a Forensic framework that is based on fog named as FoBI. This research work discussed the architecture, implementations, use cases, etc. of FoBI. This paper also provided few insights about this framework on enlightening the digital forensics processes.

Researches that have been based on misleading of justice had shown human error as a concern, which in turn leads to increased attention to cognitive bias within many forensic sectors. Thus an article [20] addressed the issues of cognitive bias. In this research, an analysis had been made upon 7 cognitive sources and discussed their countermeasures. This paper concluded that certain cognitive bias issues were same as forensic sectors, while others were different.

### 2.4 APT Attacks

Recently, people had developed a lot of interest in studying APT attack. This is due to its persistent and complex attack nature. Various researches have been conducted for preventing APT attacks. [21] article reviewed and analysed a huge number of APT attacks along with respective cases. This paper also provided an overview of APT attack types and its techniques. Usually, the APT attacks can be categorized into 5 phases based on its lifecycle [22], such as 1) reconnaissance, 2) delivery, 3) initial intrusion, 4) operation, 5) Attack benefit. Earlier attack model consisted of six phases. 1st phase is said to be intelligence gathering where the attackers spent lot of time in collecting the target information e.g. active scan, etc. 2nd phase is said to directional invasion where the attackers break into targeting network via phishing mails. Remote control is the 3rd phase where users are controlled by attackers by sending Trojan programs. Lateral movement and data mining is the 4th and 5th phases. Destruction of system or data theft is said to be the 6th phase.

A novel approach to manage cyber-attacks in fog computing is developed [23], where a cyber-tool for neutralizing the issues is adopted. The platform used here comprised of three basic entities namely provider, the attacker, and the cyber-insurer. Basically, the fog computing provider dynamically optimizes the computing resources, whereas the attacker dynamically alters the attack resource allocations in order to improve the chance of attack. In addition to that, the provider made a dynamic decision to prevent the potential loss. In this research, two levels were proposed. In the lower level, an evolutionary game is analysed with respect to cyber strategies and attacker's strategies. In the upper level, a determination strategy is utilized.

### 2.5 Shamoon Attacks

This attack is also referred as W32.DisTrack. It is discovered in the year 2012, where 32 – bit NT kernel types of MS Windows had been targeted. Due to its destructive behaviour and the massive recovery cost, this malware has been remarked to have different behaviour when compared to other malwares. [24] investigated about various cyber-attacks that are happened in Saudi Arabia. The Shamoon virus attack had been surfaced at Saudi Aramco, wiped out thousands of systems' data by replacing it with a partial image of burning American flag. [25] analyzed various security taxonomies of cyber- attacks. Here two versions of Shamoon attack had been assessed. Upon observing Shamoon; this malware targeted the energy sector, where its propagation was achieved through network sharing. This attack resulted in infection of system files and master boot record and making it inoperable.

To examine cyber – attack security, a methodology named McCumber cube is developed [26]. The primary aim is to develop a foundation for attack taxonomy. This technique utilized Shamoon attack as a case study. It is observed that, Shamoon acted as a spyware but deleted those files, once it got uploaded them to the attacker. [27] presented a case study of cyber-attacks in Saudi organizations. This paper mainly concentrates on two malwares namely Ransomware and Shamoon malware. It represented Shamoon as a wiper malware, as it wiped out hard drives. Another wiper named Shamoon 2.0 attacked various critical sectors of Saudi Arabia. This version is similar to previous version but came up with a fully operational model of a Ransomware module.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

## 3. The literature review

proposed a novel advanced persistent threat attack to analyze high volumes of network traffic effectively to divulge weak signals associated with data exfiltration and other suspect activities[29] [32]. It was one of the most significant techniques to extract serious threats in modern organizations. It could be spanned over a very long period to utilize and evade encrypted detection connections strategies depend on existing defensive solutions[37]. The advantage is that flow records could be stowed and compacted effectively even if the user gathered information over an extended period. Processing of flow records had been estimated much feasibility compared to analyzing large capacity of original traffic data. The disadvantage is that distributed denial of service became inadequate to identify advanced persistent threat attack because specialist attackers found very hard to detect a restricted number of the particular host from preventing automatic malware attacks. The final consequence had been ranked from suspicious internal hosts that permit security specialists to focus similar set of the host that are needed to be analyzed and characterized within huge organizations[37].

anticipated a machine learning correlation analysis to determine the probability of early alerts to implement an entire advanced persistent threat attack[38] [32]. The advantage is that the user could able to identify and remove cyberspace attacks easily in a systematic manner[33]. The disadvantage is that a multi-stage attack with the purpose of the cooperating system to gain information from the targeted system causes significant damage and considerable financial loss that are required to predict attacks very accurately and quickly. The capacity, complexity, and difference of cyber-attacks got increased continuously on the internet of things. A high cost had been brought out with much interest in research and development towards implementing novel cyber-attacks defense techniques. The other drawbacks are high false positive, needed important knowledge experts to maintain network attacks and detect only life cycle of advanced persistent attack but it could not attain real-time detection [38].

The Shamoon virus attack had been surfaced at Saudi Aramco wiped out with thousands of system data by substituting attack along with a partial image of the scorching American flag. Currently, the Shamoon attack [29], which is famous industrial espionage, played a massive role in destroying Saudi Aramco industrial data.There were various methods presented for detecting the Shamoon attacks[29]. [39]suggested detection of privacy threat by peculiar feature extraction malware to combat targeted Shamoon cyber-attacks. The advantage is that malware had the same objective to detect specific attacks efficiently. The refinement of artifacts in association with targeted organization technology and the expected threat had been improved with a detection probability of cyber-space attacks. The weight had been specified for every artifact based on the difference between the existence of malicious and benevolent code related to the expected organization for targeted attacks. A detection technique had been described based on peculiar artifacts that would not help in detecting considerable defense against unknown malicious attacks.

proposed a survey on fog computing to provide good computational request in real time latency sensitive applications. The advantage is that it provides superior storage facilities of cloud-based methods. The disadvantage is that congested network and high latency. The concept of fog computing had been developed to address internet of things application challenges in existing cloud computing that are shared highly combined. As the cloud datacenters were purely then user would frequently to deal with storage processing demands for different internet of things services[40].

anticipated an application for fog computing to improve better quality of experience concerning system services[34]. A fog computing delivered service application combined with less latency response requirements. Certain challenging works computing application are hierarchical, distributed and heterogeneous nature of computational instances in fog computing standards. Differentiated user had been intensified along with application placement problem. The advantage is that it provides data processing time, network congestion, resource affordability and service quality. The disadvantage is that it is required to maximize quality of experiencewith regards to utility access, resource consumption and service delivery.[29] The fog compatible instance applications had been prioritized along with several application placement request allowing user to estimate the recent status of fog computing. In fog computing environment, user had been facilitated to provide appropriate fog instances such that policy could be implemented correctly[40][41].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

proposed regarding different cyber-attacks that had been occurred in Saudi Arabia to detect attacks easily. The disadvantage is that the attack became dangerous intensification in international hacking because faceless enemies required driving an ability to impose critical physical damage that makes United States government officials, and cybersecurity researchers burden that the culprits could replicate it with other countries as several industrial plants had relied over similar American engineered computer system that had been compromised. The advantage is that it allows the user to detect attacks with much feasibility[41].

## 4. METHODOLOGY:

### 4.1 FPSO Optimizer:

The FPSO optimizer [16] impersonates the wave nature of waves by using the features of three parameters namely frequency, wavelength, and amplitude. Here the movement of the particle has been made similar to wave movement. Here, the movement of each and every particle completely depends upon the given frequency. It is noted that every particle has different frequency i.e. for every iteration, the particles not necessarily to be moved in same direction. Initially, the random position are assigned to every particle. Then the sound signal frequencies based on the velocities are updated. Depending upon the frequency, the particles are moved up and down. At every iterations, this optimizer derives a fitness functions and stores least fitness value function. Thus the FPSO optimizer improves the primary solution by converging towards the best search space point.

### 4.2 Detection of Shamoon Attack by FPSO (Investigator Digital Forensics-FPSO):

The major goal of IDF-FPSO system is to classify and detect the Shamoon attack from different attacks[42]. This system utilizes two datasets that have been used as training sets for Shamoon attack and different attack types. Initially, the trained data to feature extraction process, where 272 features are presented in each row. In this process, frequencies are determined for single and double characters. The attained frequencies are then fed into FPSO optimizer. In this optimizer, weight has been determined such that if the weighed value is greater than 0.6, then the weight is predicted. On the other hand, the features extracted from test data are fed into PSO optimizer and the weight is predicted respectively. Once the weights are determined, K-means clustering is performed for both the testing and training data. For the Shamoon type dataset and the different attack dataset, the centroid consists of two cluster classes respectively. Finally, fitness function is evaluated to determine the optimum weight. KNN classifier is utilized for classifying the training and testing data and finally predicts the sample data point.

### 4.3 System Architecture:

The overall representation of proposed methodology is depicted in figure 1. Initially, fog nodes are initiated, where the nodes delivers three types of data namely industrial, medical and educational data. Secondly Shamoon attack is created followed by distance matrix evaluation. As the Shamoon attack focuses on attacking industrial data, the attack distribution movement focuses mainly on industrial data. After the evaluation, priorities of the particles should be assigned randomly. Once FPSO parameters are initialized, objective function of every particle is evaluated. The FPSO mechanism implements the working procedure of TSP. Under the FPSO mechanism, swap and insertion operations are performed. In order to find the best shortest path, nearest neighbouring algorithm is used, which follows evaluation of fitness function. After evaluation, local best lbest and global best gbest solutions are obtained. Finally, appropriate positions and velocities are updated. This iteration will be processed, till we get the optimal solution. From the resultant optimum path, the distribution of Shamoon attack movement can be analyzed.

## 5. System Modules:

The proposed system utilizes the TSP based approach in FPSO optimizer in analysing the behaviour of Shamoon attack movement. The typical procedure for the proposed algorithm is as follows,

Step 1: Initialization
Step 2: Calculating fitness value
Step 3: Start iteration
Step 4: Swap and insert operation
Step 5: Update pbest and gbest
Step 6: Update the position of the particle
Step 7: Judge the terminating criteria
Step 8: Output the optimal path

### 5.1 Initialization:

#### 5.1.1 Initializing Fog Nodes:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Usually fog computing can also be termed as 'fogging' in which a smart device handles specific application processes at the edge of the network, but remaining are managed in cloud environment [28]. It is an intermediate layer between the cloud and the device and provides efficiency in analysing and storing the data. This computing has its own advantages like location awareness and minimum latency, wider geographical distribution and preferable supports mobility.

It is comprised of large number of nodes. The primary task of fog is delivering data to the users who are located at the edge of the network. The term edge signifies various nodes where the end user is connected. In this module, fog nodes are initiated where three types of datasets are employed. These three datasets comprises features from medical data *(M)*, industrial data *(I)*, and educational data *(E)*. Each data has its own priority and cost.

### 5.2 Creation of Shamoon Attack:

The Shamoon virus [3] is one of the most sophisticated malware types. An essential element in latest version of Shamoon is embedded user credentials that are obtained from previous attacks. Here, this technique is as simpler as hackers obtaining passwords or a long term attack which involves phishing of passwords or an inside job. This malware consists of three separate phases namely dropper, wiper, and reporter. In dropper phase, the virus got attached into the system and spread across its local network making copies of it and finally drop certain components. In wiper phase, which is also known as Disstrack, the malware uses the currently existing EldoS software to override the hard disk data and MBR. The final phase is the reporter phase, where the communication is handled by the hacker by controlling the server.

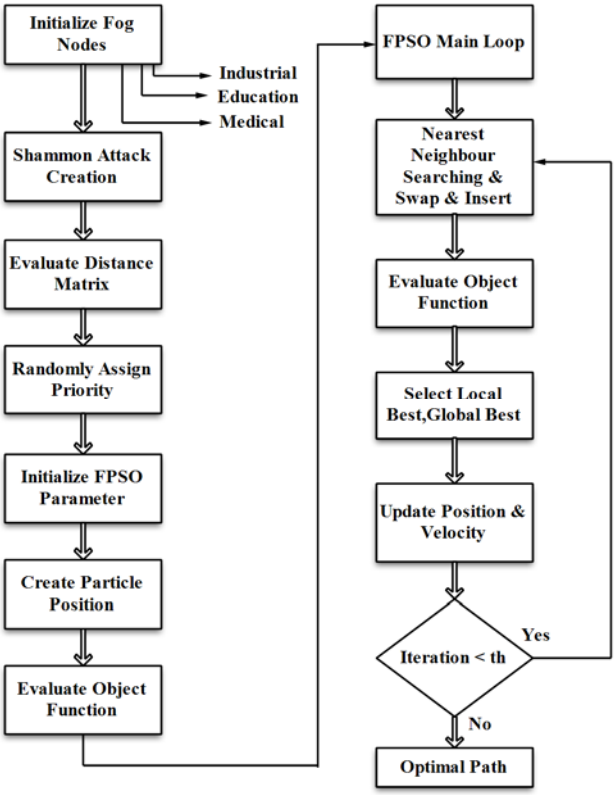


Figure 1: Block Diagram Representation of Proposed Methodology.

In this proposed system, we are focusing on analysing the behaviour of Shamoon attack. Thus the priority should be given to industrial data *(I)*, so that the Shamoon will move towards it. After providing high priority to the Shamoon data. After initialization of number of nodes and creation of Shamoon (as particles), the distance matrix $d_m$ has to be estimated. It can be formulated as below,

$$d_m = \sum_{i=1}^{n-1} d(p_i, p_{i+1}) + d(p_n, p_1) \quad (1)$$

Where $d(p_i, p_{i+1})$ represents the distance between particles$p_i, and\ p_{i+1}$. The best path will contain the path with the shortest distance between subsequent particles. Thus the distance is calculated between the particles and save them to the distance matrix $d_m$.

### 5.3 Distance matrix

In this proposed system; we are focusing on analyzing the behavior of Shamoon attack. Thus the priority should be given to industrial data (I), so that the Shamoon will move towards it. After providing high priority to the Shamoon data, and after initialization of number of nodes and creation of Shamoon (as particles), the distance

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

matrix d_m has to be estimated. It can be formulated in presented equation 2.

$$d_m = \sum_{i=1}^{n-1} d(p_i, p_{i+1}) + d(p_n, p_1) \qquad (2)$$

Where d( p_i,p_(i+1) ) represent the distance between particlesp_i, 〖and p〗 _(i+1). The best path will contain the path with the shortest distance between subsequent particles. Thus, the distance is calculated between the particles and save them to the distance matrix 〖 d〗 _m.

### 5.4 Randomly assign prioritization

FOG nodes are initiated where three types of datasets are employed. These three datasets comprise features from medical data (M), industrial data (I), and educational data (E). Each data has its own priority and cost. Shamoonattack target the industrial data so that it get the highest priority to identify Shamoon attack.

### 5.5 Initializing FPSO Parameter

FPSO learning [16] consists of two types of dataset namely Shamoon attack data and another attack data. This data has been classified based on feature selection and weight. Initially, we have to find the encoded variables,

$$0 \leq W \leq 1$$

If the value of the weight W > 0.6, then those features are selected and remaining are neglected. The feature size is reduced as below,

$$X_1 = W.X$$

where$X$ represents training data features, $W$ represents value of the weight, $X_1$represents new feature value.

$$0 \leq W \leq 1^{min}(A/B)$$

### 5.6 Create particle position

After the Initializing FPSO Parameter each particle is given a random position in the FOG computing environment. And this determines the best path

### 5.7 Fitness value Evaluation:

Fitness / object function is used for assessing the quality of the solution. The function's input denotes the class of the selected features. In FPSO, KNN classifier is used for classification purpose. KNN classifier can be built using the selected features. Thus the fitness function for classifiers A and B are calculated which is formulated as below,

$$Fitness = \left(\frac{A}{B}\right) \qquad (3)$$

### 5.8 Nearest Neighbor Searching:

As mentioned in proposed system, this research investigates local nearest searching techniques to improve the quality of the solutions. Initially, every data nodes are assigned with different priority level. But our aim is to provide high priority level to industrial data, as Shamoon concentrate on industrial data. By using priority based sorting and by utilizing the fitness value obtained from FPSO optimizer, the nearest neighbor is searched and the particles with minimum distance has been estimated. This can be done by inserting and swapping operations.

#### 5.8.1 Inserting & Swapping:

The operators namely SWAP and INSERT are implemented to the copy of every solutions for every iteration in order to generate neighbouring solutions and also to enhance the search performance of the neighborhood solutions. These to operators are described below,

**INSERT:**

When considering insertion, the set U $p_i$has all the features that are selected. As the selected features are important, this would be possible in order to improve the classification accuracy by adding one feature from the specified set to the present set of selected features.

**SWAP:**

When compared with Removing / inserting operator, swapping is a complicated one. For every solution, the operation will act on both the selected and unselected sets. Instead of adding or removing a feature, this swapping operation will find the better solutions by swapping in between the features. The working of insert and swap operator are depicted in figure 2.

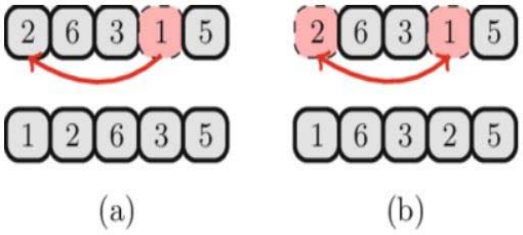


*Figure 2: Representation of Inserting (a) and Swapping (b)*

From the above operations pbest and gbest solutions are updated. Thus the new particle

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

position is updated. The pseudo code of the proposed system has been explained below.

**Pseudo Code**

1) Initialization

   Initialize Fog nodes

   Initialize FPSO Parameters

2) Priority Assign

   P = Selected nodes (nFog)

   Pr = random (P)

3) Evaluation

   FPSO Pseudo Code

   // Initialize

   For i =1 : nPop

   P → Random Particle Priority

   V → Initialize Velocity

   End

   // main loop

   a) For i = 1:maximum Iteration

      P[I:J] = P[J:I]

      //swap & Insert

      Pbest→ Nearest Neighbour (P);

      Select local best;

      Select global best; //Store value

      $P_{new}$= f (Pold, weight);

      $V_{new}$ = max (V) || min(V);

      end

      //plot result

Display value & index;

Thus by tracing the very shortest route using TSP we could analyse that certain nodes with loop holes are exploited by the external attacker to attack the industrial data. Using the proposed approach, the source of the attack and the information of the attacker can be determined by locating the first node that got attacked. Usually, there are several methods which are employed in recovering the data. Once the data got recovered. From the recovered data, possible evidences present within the data are collected for locating the attacker using the proposed research analysis. With the help of the collected evidences, future attacks can also be prevented. This is the primary aim of this research.

### 5.9 Evaluation objective function and selection of local best and global best

Figure 2 represent operations of the swapping and inserting and this help us to evaluate the pbest and gbest solutions to obtain the best gbest point. By updating best pbest and gbest**.**

### 5.10 Update the position and velocity

In this step the particle position and velocity is updated at each iteration during the process of detecting Shamoon attack and this will help to track the shortest path.

## 6 Experiments and Results

### 6.1 Attack Distribution:

The Shamoon attack distribution has been represented graphically in below figure 3. Here three types of data are taken into consideration namely industrial, medical and educational. Here, industrial data is represented in red Color, education data in blue, medical data in yellow.

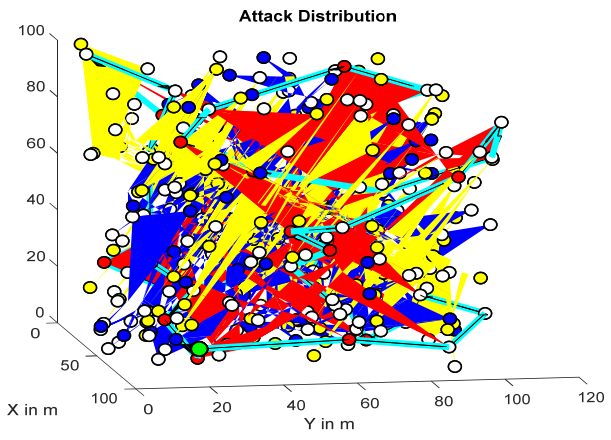


*Figure 3: Distribution of Attack Behaviour in Fog*

From the figure 3, it has been observed that the 3D- model of the Fog environment provides an extension to the edge of the network. As previously mentioned, there are three types of data namely medical, industrial and educational data are used. Shamoon data is considered to be an industrial espionage. So it will move towards the industrial data. Thus the industrial data should

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

be given high priority and cost that other two data.

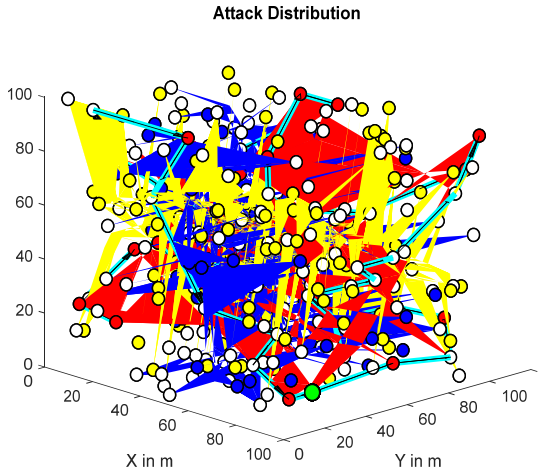


*Figure 4: Distribution of Attack Behaviour in Fog Environment*

The distribution of attack behaviour in fog is represented in 3D graphically using figure 4. The above figure shows the details of 3D model fog, 3D routing path, and the attack distribution path. A gateway which is served as fog nod can be compromised or replaced by a fake node.

### 6.2 Best Cost Estimation:

By using this distributed FPSO algorithm, the best cost can be identified. It can be represented graphically using figure 5.

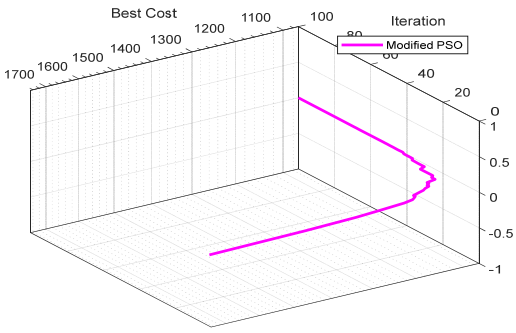


*Figure 5: Best Cost Estimation*

Here, it is considered to be over 200 iterations. For each iteration, the best cost has been computed. This cost is related to Shamoon attack data distribution.

### 6.3 Fitness Value Estimation:

Here the fog nodes are considered to be over 100. For each and every iteration, distance cost has been calculated. For every iteration, distance cost decreases. The average fitness for every iteration can be expressed as below,

Average fitness = (fitness of each particle / Total number of particles)

The result of fitness valuation has been represented graphically using figure 6.

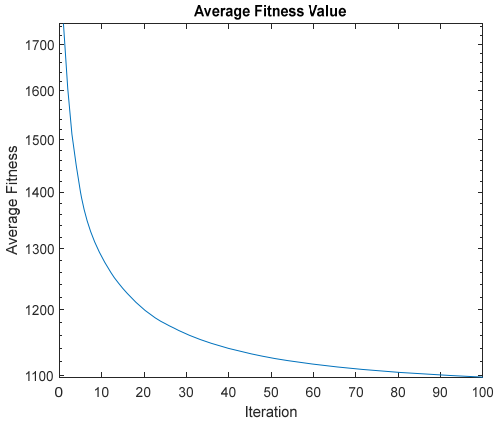


*Figure 6: Average Fitness Value Estimation*

### 6.4 Trajectory Value Evaluation:

Initially, the 1st variable is plotted based on every iteration. In FPSO, 100 nodes has been selected, where for every node the priority value got changed in each iteration. This can be represented graphically using figure 7.

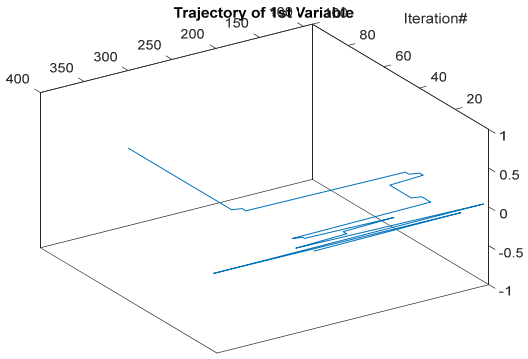


*Figure 7: Trajectory Value Estimation*

-

### 6.5 Search History Evaluation:

Upon considering the feature extraction, features are searched based on search history and the new features have been selected based on

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

weight values. This is represented graphically using figure 8.

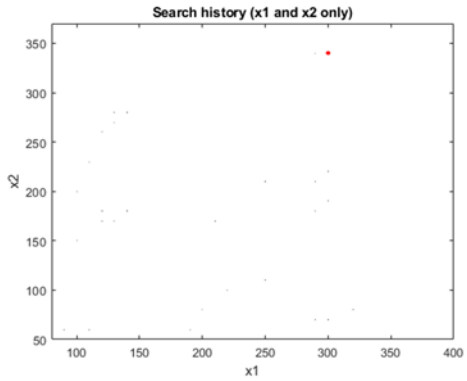


*Figure 8: Search History Estimation*

### 6.6 Location:

In order to locate the Shamoon attack, we have to focus on industrial data, rather than medical data and educational data. As the Shamoon is the industrial espionage, it will locate tor move towards the industrial data. Figure 9, plots the movement of Shamoon attack which is represented in blue colour graphically. In this graph, the infected nodes represent the industrial data and the marked path refers to the Shamoon attack movement. The movement of the attack is forward directional. The attack will search for nodes with shorter distance and moved accordingly. For example, from the graph, when considering node 26, the attack searches for shortest distance node and transmitted to node 28, instead of node 11.

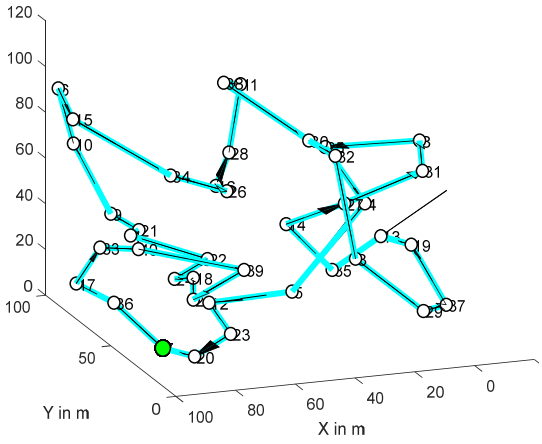


*Figure 9: Movement Path of Infection Node Attacks*

From the figure 9, the attack path movement has been clearly mentioned. Here the green node indicates the starting path and the arrow indicates the path movement. From the below figure, the attack is minimal in middle level and high in starting and ending position. From this result, it has been inferred that Shamoon attack in Fog computing targets the industrial data, weaker nodes and spreads across the server by tracing very shortest route available based on Travelling Salesman Problem (TSP).Some nodes consists of a loophole which is exploited by the external attacker to attack the industrial data. So the source of the attack is determined by locating the first node which is attacked by malware. This node is represented by green Color. On finding the first node to be attacked, the source of the attacker can be tracked. The industrial data lost can be recovered by following some techniques. Then in the recovered data, possible evidences present within the data are collected for locating the attacker. With the help of the collected evidences, future attacks can also be prevented.

## 7 CONCUSSIONS

In this research, by tracing the very shortest route using TSP, certain nodes with loop holes that are exploited by the external attacker to attack the industrial data analysed. used the proposed approach, the source of the attack and the information of the attacker is determined by locating the first node that got attacked. Usually, there are several methods which are employed in recovering the data. Once the data got recovered. From the recovered data, possible evidences present within the data are collected for locating the attacker using this method. With the help of the collected evidences, future attacks can also be prevented. For this, analysed for Shamoon attack behaviour using FPSO optimizer based on TSP studied and discussed. Initially fog nodes have been initiated with three types of data namely industrial, medical and educational data. From the provided particles, distance matrix evaluated and stored. Once FPSO parameters are initialized, objective function of every particle is evaluated. The FPSO mechanism implements the working procedure of TSP. Under the FPSO mechanism, swap and insertion operations are performed. In order to find the best shortest path, nearest neighbouring algorithm (KNN) is used, which follows evaluation of fitness function. After evaluation, local best lbest and global best gbest solutions are obtained. Finally, appropriate positions and velocities are updated. From the resultant optimum path, the distribution of Shamoon attack movement analysed. The performance of this proposed system has been

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

evaluated by estimating the fitness value, best cost. The attack distribution of Shamoon data has been observed. Using this research, we collected the digital evidence of the Shamoon attack behaviour even in the environment containing huge amounts of data. Thus an easy tracking of infected sites in Fog environment attained.

A threat intelligence scheme is proposed for the investigating and analysis behaviour and spread of Shamoon attacks in edges of Fog computing.

The type of infection that affects the data should be determined and the spreading behavior of Shamoon attack, detection path and industrial espionage .

**REFERENCES:**

[1] S. Khan, S. Parkinson, and Y. Qin, "Fog computing security: a review of current applications and security solutions," *Journal of Cloud Computing,* vol. 6, p. 19, 2017.

[2] S. Zhioua, "The middle east under malware attack dissecting cyber weapons," in *2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops*, 2013, pp. 11-16.

[3] D. Tarakanov, "Shamoon the Wiper in details," *Kaspersky Labs,* 2012.

[4] E. F. G. Goldbarg, G. R. de Souza, and M. C. Goldbarg, "Particle swarm for the traveling salesman problem," in *European Conference on Evolutionary Computation in Combinatorial Optimization*, 2006, pp. 99-110.

[5] E. Khalil, H. Dai, Y. Zhang, B. Dilkina, and L. Song, "Learning combinatorial optimization algorithms over graphs," in *Advances in Neural Information Processing Systems*, 2017, pp. 6348-6358.

[6] S. S. Juneja, P. Saraswat, K. Singh, J. Sharma, R. Majumdar, and S. Chowdhary, "Travelling Salesman Problem Optimization Using Genetic Algorithm," in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, 2019, pp. 264-268.

[7] M. Mahi, O. K. Baykan, and H. Kodaz, "A new approach based on particle swarm optimization algorithm for solving data allocation problem," *Applied Soft Computing,* vol. 62, pp. 571-578, 2018.

[8] T. Dokeroglu, E. Sevinc, and A. Cosar, "Artificial bee colony optimization for the quadratic assignment problem," *Applied Soft Computing,* vol. 76, pp. 595-606, 2019.

[9] A. A. Yahya, "Swarm intelligence-based approach for educational data classification," *Journal of King Saud University-Computer and Information Sciences,* 2017.

[10] L. Brezočnik, "Feature selection for classification using particle swarm optimization," in *IEEE EUROCON 2017-17th International Conference on Smart Technologies*, 2017, pp. 966-971.

[11] Z. Li, Q. Chen, and V. Koltun, "Combinatorial optimization with graph convolutional networks and guided tree search," in *Advances in Neural Information Processing Systems*, 2018, pp. 539-548.

[12] A. H. Halim and I. Ismail, "Combinatorial optimization: comparison of heuristic algorithms in travelling salesman problem," *Archives of Computational Methods in Engineering,* vol. 26, pp. 367-380, 2019.

[13] M. Verma and A. Chauhan, "5/4 approximation for Symmetric TSP," *arXiv preprint arXiv:1905.05291,* 2019.

[14] V. Raman and N. S. Gill, "Review of different heuristic algorithms for solving Travelling Salesman Problem," *International Journal of Advanced Research in Computer Science,* vol. 8, 2017.

[15] M. Mulani and V. L. Desai, "Computational Performance Analysis of Ant Colony Optimization Algorithms for Travelling Sales Person Problem," in *Proceedings of International Conference on ICT for Sustainable Development*, 2016, pp. 561-569.

[16] M. S. a. a. R. Hwaitat A. , "AN ENHANCED PARTICLE SWARM OPTIMIZATION USING FREQUENCIES WAVE SOUND (FPSO)," *Journal of Theoretical and Applied Information Technology,* vol. 97, 2019.

[17] P. More, 1, J. K. , and 2, "Fog Computing," *International Research Journal of Engineering and Technology*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*(IRJET),* vol. 04, pp. P1113 - P1116, 2017.

[18] T. Wang, J. Zeng, Y. Lai, Y. Cai, H. Tian, Y. Chen, *et al.*, "Data collection from WSNs to the cloud based on mobile Fog elements," *Future Generation Computer Systems,* 2017.

[19] E. Al-Masri, Y. Bai, and J. Li, "A Fog-Based Digital Forensics Investigation Framework for IoT Systems," in *2018 IEEE International Conference on Smart Cloud (SmartCloud)*, 2018, pp. 196-201.

[20] N. Sunde and I. E. Dror, "Cognitive and human factors in digital forensics: Problems, challenges, and the way forward," *Digital Investigation,* vol. 29, pp. 101-108, 2019.

[21] M. Li, W. Huang, Y. Wang, W. Fan, and J. Li, "The study of APT attack stage model," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, 2016, pp. 1-5.

[22] W. Niu, X. Zhang, G. Yang, R. Chen, and D. Wang, "Modeling attack process of advanced persistent threat using network evolution," *IEICE TRANSACTIONS on Information and Systems,* vol. 100, pp. 2275-2286, 2017.

[23] S. Feng, Z. Xiong, D. Niyato, P. Wang, and A. Leshem, "Evolving Risk Management Against Advanced Persistent Threats in Fog Computing," in *2018 IEEE 7th International Conference on Cloud Networking (CloudNet)*, 2018, pp. 1-6.

[24] N. Perlroth and C. Krauss, "A cyberattack in Saudi Arabia had a deadly goal. Experts fear another try," *The New York Times,* vol. 15, 2018.

[25] R. Derbyshire, B. Green, D. Prince, A. Mauthe, and D. Hutchison, "An Analysis of Cyber Security Attack Taxonomies," in *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2018, pp. 153-161.

[26] C. Easttom and W. Butler, "A Modified McCumber Cube as a Basis for a Taxonomy of Cyber Attacks," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, 2019, pp. 0943-0949.

[27] S. Alelyani and H. Kumar, "Overview of Cyberattack on Saudi Organizations," 2018.

[28] S. Yi, C. Li, and Q. Li, "A survey of fog computing: concepts, applications and issues," in *Proceedings of the 2015 workshop on mobile big data*, 2015, pp. 37-42.

[29] Al Hwaitat, S. Manaseer, &R. Al-Sayyed,. (2019). A Survey of Digital Forensic Methods under Advanced Persistent Threat in Fog Environment, Journal of Theoretical and Applied Information Technology , Vol.97. No 18,PP. 4934-4954. 26.

[30] A. Al Hwaitat, S. Manaseer ,(2018), Centralized Web Application Firewall Security System, Modern Applied Science; Vol. 12, No. 10; PP.164-170. https://doi.org/10.5539/mas.v12n10p16427.

[31] A. Al Hwaitat, S. Manaseer and R. Jabri ,(2018) ,Distributed Detection and prevention of Web Threats in Heterogeneous Environment , Modern Applied Science; Vol. 12 ,No10 ,PP.13-22. https://doi.org/10.5539/mas.v12n10p1328.

[32] A. Al Hwaitat, S. Manaseer ,(2017), Validation and Integrity Mechanism for Web Application Security, International Journal of Engineering Research & Science , Vol. 3, No.11,PP.34-38. 29.

[33] O. rababha , A. Al Hwaitat , S. Manasser ,(2016) Web Threats Detection and Prevention Framework, communications and Network, Vol. 8, No.8, PP. 170- 178. https://doi.org/10.4236/cn.2016.83017.

[34] A. Al Hwaitat , M. Qasem ,R. Fabozzi ,(2020)," Security of Data Access in Fog Computing using Location-based Authentication", International Journal of Advanced Trends in Computer Science and Engineering ,Vol. 9,No. 1 ,PP.247-253.

[35] A. Al Hwaitat , M. Qasem ,(2020)," A Survey on Li Fi Technology and Internet of Things (IOT)", International Journal of Advanced Trends in Computer Science and Engineering ,Vol. 9,No. 1 ,PP.225-253

[36] A. Hudaib , A. Al Hwaitat ,(2018), " Movement Particle Swarm Optimization Algorithm" Modern Applied Science; Vol. 12, No. 1,pp.148-164.

[37] M., Marchetti, F. Pierazz,, M., Colajanni, and A. Guidoet, (2016). "Analysis of high volumes of network traffic for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

advanced persistent threat detection." Computer Networks109: 127-141

[38] I. Ghafir, M. Hammoudeh, V. Prenosil,H.Robert, and R. Khaled, (2018). "Detection of advanced persistent threat using machine-learning correlation analysis." Future Generation Computer Systems89: 349-359.

[39] F.Ahmad, K. Batool, and A. Javed, (2016). Detection of Privacy Threat by Peculiar Feature Extraction in Malwares to Combat Targeted Cyber Attacks. Advanced Computer and Communication Engineering Technology, Springer: 1237-1247.

[40] R. Mahmud, S. Srirama , K., Ramamohanarao, and R. Buyya, (2019). "Quality of Experience (QoE)-aware placement of applications in Fog computing environments." Journal of Parallel and Distributed Computing132: 190-203.

[41] N. Perlroth, and C. Krauss (2018). "A cyberattack in Saudi Arabia had a deadly goal. Experts fear another try." The New York Times15.

[42] A. Al Hwaitat, S. Manaseer, R. Al-sayyed3 ,m almaiah, o almomani ,(2020) , An Investigator Digital Forensics Frequencies Particle Swarm Optimization For Dectection And Classification Of Apt Attack In Fog Computing Enviroment (IDF-FPSO) , Journal of Theoretical and Applied Information Technology, Vol.98 Issues 07.